Multivariate Anomaly Detection in Mixed Telemetry time-series Using A Sparse Decomposition

Barbara Pilastre, Jean-Yves Tourneret
TéSA/IRIT/INP-ENSEEIHT
Toulouse, France
Email: barbara.pilastre@tesa.prd.fr
Email: jean-yves.tourneret@enseeiht.fr

Stéphane D'Escrivan
Centre National d'Etudes Spatiales (CNES)
Toulouse, France
Email: Stephane.d-Escrivan@cnes.fr

Loïc Boussouf
Airbus Defence and Space
Toulouse, France
Email: loic.boussouf@airbus.com

***Abstract*—Spacecraft health monitoring from housekeeping telemetry data represents one of the main issues in space operations. Motivated by the success of machine learning or data driven-based methods in many signal and image processing applications, some of these methods have been applied to anomaly detection in housekeeping telemetry via a semi-supervised learning. This paper studies a new multivariate anomaly detection algorithm based on a sparse decomposition on a dictionary of nominal patterns. One originality of the proposed method is a multivariate framework allowing us to take into account possible relationships between different telemetry parameters, in particular through a joint processing of time-series described by mixed continuous and discrete parameters. The proposed method is tested with real satellite telemetry and evaluated on a representative anomaly dataset composed of actual anomalies that occurred on several operated satellites. The first results confirm the interest of the proposed method and demonstrate its competitiveness with respect to the state-of-the-art.**

I. Introduction

A main issue in space operations is to ensure the proper conduct of the missions by monitoring spacecraft health and detecting failures as soon as possible. Spacecraft health monitoring is classically performed by monitoring telemetry times series using anomaly detection (AD) techniques [1], [2]. Housekeeping telemetry data consists of hundred to thousand telemetry parameters corresponding to various sensors. Some of these parameters (such as antenna positions or equipment operating mode ON/OFF) take few values and can thus be considered as observations of discrete random variables (after a possible reparametrisation, e.g., "ON= 0" and "OFF= 1". The other parameters (such as temperature, pressure, voltage etc...) are observations of continuous variables. As a consequence, the whole set of telemetry data can be considered as a multivariate time-series of mixed discrete and continuous data.

Motivated by the success of machine learning (ML) or data driven-based methods in signal and image processing, some AD methods have been investigated for housekeeping telemetry via a semi-supervised learning. In a first step referred to as learning, the algorithm builds a reference model from past telemetry data describing only nominal (also called normal) operation of the spacecraft. In a second step referred to as detection, an appropriate comparison to the reference model allows the detection of potential anomalies affecting most recent data. ML-based methods for AD in housekeeping telemetry can be divided in two categories depending on the kind of data processed by the algorithms, i.e., univariate or multivariate time-series. Univariate AD methods handle each telemetry parameter independently and detect univariate anomalies corresponding to abnormal behaviour (never seen before) of an individual parameter. Examples of univariate anomalies are displayed in Fig. 1 (see examples 1, 2, 3, 5 and 6 in the red boxes). The ML-based techniques investigated in this framework are based on powerful algorithms that include the one-class support vector machine [3], nearest neighbour techniques [4]–[6] or neural networks [7]. These solutions allowed spacecraft health monitoring to be significantly improved. However, by definition these methods do not take into account possible relationships between different parameters, which makes it difficult (say impossible) to detect multivariate or contextual anomalies corresponding to changes in these relationships. Examples of multivariate anomalies affecting real telemetry data are displayed in Fig. 1 (examples 4 and 7 in the red boxes). Detecting anomalies in these time-series requires to handle several telemetry time-series jointly in order to facilitate the detection of multivariate anomalies. This framework has been recently investigated in [8]–[10].

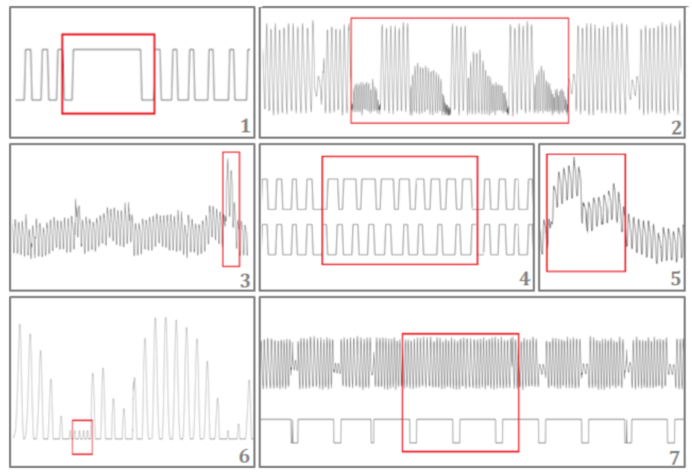

Fig. 1. Examples of univariate and multivariate anomalies (blue boxes)

This paper studies a new multivariate AD method for mixed telemetry data based on a sparse decomposition on a dictionary of nominal patterns. Inspired by [11], the idea of the proposed method is to build a dictionary of nominal patterns and to decompose most recent data into this dictionary in order to detect potential anomalies from residues of the sparse decomposition. The paper is organized as follows.

Section II introduces the proposed AD method for mixed telemetry data based on a sparse decomposition. Section III evaluates its performance via a comparison with three state-of-the-art methods that are tested on a representative dataset with available ground-truth. Conclusion and future works are reported in Section IV.

II. Detecting anomalies in mixed data using a sparse decomposition

This section focuses on the detection step and assumes that the dictionary of nominal patterns has been learned using past telemetry describing only normal operation of the spacecraft. Learning a dictionary with discrete and continuous atoms is an interesting and challenging problem which will be considered in future work.

A. Preprocessing

The preprocessing is divided into two steps: 1) segmenting the telemetry times-series into overlapping windows of fixed size w with a shift δ as illustrated in Fig. 2, and 2) vectorizing the resulting matrices by concatenating their lines. A mixed vector $\boldsymbol{y} \in \mathbb{R}^N$ obtained after this preprocessing is divided into K blocks, i.e., $\boldsymbol{y} = [\boldsymbol{y}_1^T, ..., \boldsymbol{y}_K^T]^T$ where $\boldsymbol{y}_k \in \mathbb{R}^w, k = 1, ..., K$ is the kth block of $\boldsymbol{y}$ associated with the kth parameter, K is the number of telemetry parameters and w is the number of samples of the time window (sample size).

B. Anomaly Detection Using a Sparse Decomposition

Recent years have witnessed a growing interest for sparse decomposition in many signal and image processing applications [12]–[14] and especially for AD [11], [15], [16]. Inspired by [11], the proposed AD strategy decomposes a test signal with mixed components as a sum of 1) a nominal signal expressed as a linear combination of few columns (called atoms) of a known dictionary, 2) a possible anomaly signal and 3) an additive noise such as

$$\boldsymbol{y} = \boldsymbol{\Phi x} + \boldsymbol{e} + \boldsymbol{b} \tag{1}$$

where $\boldsymbol{y} \in \mathbb{R}^N$ is the test signal, $\boldsymbol{\Phi} \in \mathbb{R}^{N \times 2L}$ is a block-diagonal dictionary of nominal patterns (see (2) for detail) previously learned using past normal telemetry data, $\boldsymbol{x} \in \mathbb{R}^{2L}$ is a sparse vector of coefficients, $\boldsymbol{e} \in \mathbb{R}^N$ is an anomaly signal ($\boldsymbol{e} = \boldsymbol{0}$ in absence of anomaly and $\boldsymbol{e} \neq \boldsymbol{0}$ in presence of anomaly) and $\boldsymbol{b} \in \mathbb{R}^N$ is an additive noise. Note that the anomaly detection can be formalised as a hypothesis test on vector $\boldsymbol{e}$ (Hypothesis H0: $\boldsymbol{e} = 0$ against alternative H1: $\boldsymbol{e} \neq 0$). The originality of the proposed method lies in handling mixed discrete and continuous telemetry time-series in order to take into account relationships between parameters and to detect univariate and multivariate anomalies. To this end, the proposed method considers a block diagonal dictionary, with two blocks denoted as $\boldsymbol{\Phi}_\text{D} \in \mathbb{R}^{N_\text{D} \times L}$ and $\boldsymbol{\Phi}_\text{C} \in \mathbb{R}^{N_\text{C} \times L}$ respectively composed of discrete and continuous atoms, and applies two distinct strategies depending on whether data are discrete or continuous. Note that the first block $\boldsymbol{\Phi}_\text{D}$ (dictionary of discrete nominal patterns) and the second block $\boldsymbol{\Phi}_\text{C}$ (dictionary of continuous nominal patterns) have been learned jointly in order to take into account possible correlations between the different time-series, especially between discrete and continuous ones. Note that the lth discrete atom of the discrete dictionary Φ_D

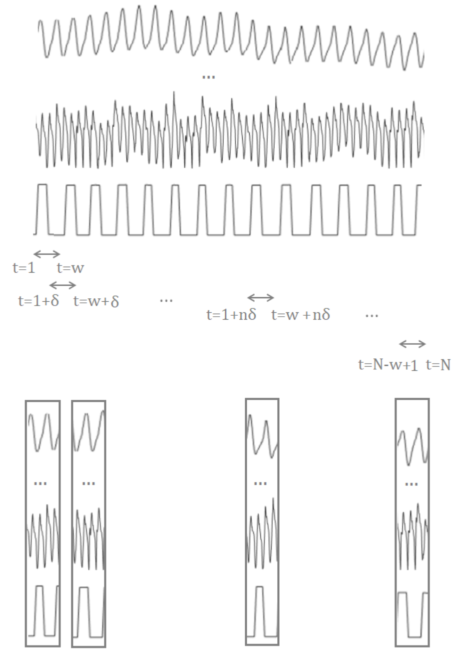

Fig. 2. Segmentation of telemetry time-series into overlapping windows

and the lth continuous atom of the continuous dictionary Φ_C are composed of discrete and continuous behaviours observed in the same multivariate atom. The signals $\boldsymbol{y}$, $\boldsymbol{x}$, $\boldsymbol{e}$ and $\boldsymbol{b}$ are divided into discrete and continuous counterparts denoted as $\boldsymbol{y}_\text{D}, \boldsymbol{e}_\text{D}, \boldsymbol{b}_\text{D} \in \mathbb{R}^{N_\text{D}}$ and $\boldsymbol{x}_\text{D} \in \mathbb{R}^L$, and $\boldsymbol{y}_\text{C}, \boldsymbol{e}_\text{C}, \boldsymbol{b}_\text{C} \in \mathbb{R}^{N_\text{C}}$ and $\boldsymbol{x}_\text{C} \in \mathbb{R}^L$, leading to the following decomposition

$$\begin{bmatrix} \boldsymbol{y}_\text{D} \\ \boldsymbol{y}_\text{C} \end{bmatrix} = \begin{bmatrix} \boldsymbol{\Phi}_\text{D} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{\Phi}_\text{C} \end{bmatrix} \begin{bmatrix} \boldsymbol{x}_\text{D} \\ \boldsymbol{x}_\text{C} \end{bmatrix} + \begin{bmatrix} \boldsymbol{e}_\text{D} \\ \boldsymbol{e}_\text{C} \end{bmatrix} + \begin{bmatrix} \boldsymbol{b}_\text{D} \\ \boldsymbol{b}_\text{C} \end{bmatrix}. \tag{2}$$

The proposed method processes discrete data in a first step, by estimating the vector $\boldsymbol{x}_\text{D}$ of the discrete sparse decomposition and the discrete anomaly signal $\boldsymbol{e}_\text{D}$. The output of this first step is then used to select the atoms preserving the existing relationships between discrete and continuous parameters and detect multivariate anomalies. The continuous approximation is performed in a second step by using the selected atoms. Given the applied preprocessing, the anomaly signals $\boldsymbol{e}_\text{D}$ and $\boldsymbol{e}_\text{C}$ are respectively divided into K_D and K_C groups corresponding to the numbers of discrete and continuous parameters, i.e., $\boldsymbol{e}_\text{D} = [\boldsymbol{e}_{\text{D},1}^T, ..., \boldsymbol{e}_{\text{D},K_\text{D}}^T]^T$ and $\boldsymbol{e}_\text{C} = [\boldsymbol{e}_{\text{C},1}^T, ..., \boldsymbol{e}_{\text{C},K_\text{C}}^T]^T$. The discrete and continuous sparse representations are described below.

1) Discrete Sparse Decomposition: In order to detect potential anomalies in the discrete signal $\boldsymbol{y}_\text{D}$ we propose to solve the following problem

$$\arg \min_{\boldsymbol{x}_\text{D} \in \mathcal{B}, \boldsymbol{e}_\text{D} \in \mathbb{R}^{N_\text{D}}} \| \boldsymbol{y}_\text{D} - \boldsymbol{\Phi}_\text{D} \boldsymbol{x}_\text{D} - \boldsymbol{e}_\text{D} \|_2^2 + b_\text{D} \sum_{k=1}^{K_\text{D}} \|\boldsymbol{e}_{\text{D,k}}\|_2 \tag{3}$$

where $\|\mathbf{e}_{\text{D},k}\|_2, k = 1, ..., K_\text{D}$ is the Euclidean norm, $\mathbf{e}_\text{D}, k$ corresponds to the kth time-series of $\mathbf{e}_\text{D}$ associated with the kth discrete parameter and b_D is a regularization parameter that controls the level of sparsity of $\boldsymbol{e}_\text{D}$. The model imposes two distinct sparsity for $\boldsymbol{x}_\text{D}$ and $\boldsymbol{e}_\text{D}$. The sparsity of $\boldsymbol{e}_\text{D}$ is ensured by the last term and reflects the fact that anomalies are rare and affect few parameters at the same time. The vector $\boldsymbol{x}_\text{D}$ is constrained to belong to $\mathcal{B}$, where $\mathcal{B}$ is the canonical or natural basis of $\mathbb{R}^L$, i.e., $\mathcal{B} = \{\boldsymbol{\epsilon}_l, l = 1, \cdots, L\}$, where $\boldsymbol{\epsilon}_l$ is the lth canonical vector (whose components equal 0 except its lth one equal to 1). The discrete sparse decomposition is a combinatorial problem which solves for each discrete atom $\boldsymbol{\phi}_{\mathrm{D},l}, l = 1, ..., L$ (with L the number of atoms in Φ_{D}) the following problem

$$\widehat{\mathbf{e}}_{\mathrm{D},l} = \arg\min_{\mathbf{e}_{\mathrm{D},l}} \|\boldsymbol{y}_{\mathrm{D}} - \boldsymbol{\phi}_{\mathrm{D},l} - \boldsymbol{e}_{\mathrm{D},l}\|_2^2 + b_{\mathrm{D}} \sum_{k=1}^{K_{\mathrm{D}}} \|\mathbf{e}_{\mathrm{D,k}}\|_2. \quad (4)$$

The solution of the optimization problem (4) is classically obtained using the shrinkage operator [17] $\widehat{\mathbf{e}}_{\mathrm{D},l} = T_{b_{\mathrm{D}}}(\mathbf{h}_{\mathrm{D}})$ with

$$[T_{b_{\mathrm{D}}}(\mathbf{h})]_{\mathrm{k}} = \begin{cases} \frac{\|\mathbf{h}_{\mathrm{k}}\|_2 - b_{\mathrm{D}}}{\|\mathbf{h}_{\mathrm{k}}\|_2} \mathbf{h}_{\mathrm{k}} & \text{if } \|\mathbf{h}_{\mathrm{k}}\|_2 > b_{\mathrm{D}} \\ 0 & \text{otherwise} \end{cases} \quad (5)$$

where $\mathbf{h}_{\mathrm{D}} = \boldsymbol{y}_{\mathrm{D}} - \boldsymbol{\phi}_{\mathrm{D},l}$ $\mathbf{h}_k$ is the kth part of $\mathbf{h}_{\mathrm{D}}$ associated with the kth time-series for $\mathrm{k} = 1, ..., K_{\mathrm{D}}$. The discrete anomaly detection looks for the anomaly vectors $\widehat{\boldsymbol{e}}_{\mathrm{D},l}, l = 1, ..., L$ resulting from the discrete sparse decomposition that are equal to 0 and builds a subset $\mathcal{M}$ defined as

$$\mathcal{M} = \{l \in \{1, \cdots, L\} | \ \|\widehat{\mathbf{e}}_{\mathrm{D},l}\|_2 = 0\}. \quad (6)$$

The subset $\mathcal{M}$ contains the values of l associated with the discrete atoms $\phi_{\mathrm{D},l}$ that are the closest to $\boldsymbol{y}_{\mathrm{D}}$. An anomaly on the discrete signal is declared if $\mathcal{M}$ is empty.

2) Continuous Sparse Decomposition: In order to detect potential anomalies in the continuous signal $\boldsymbol{y}_{\mathrm{C}}$ we propose to solve the following problem

$$\min_{\boldsymbol{x}_{\mathrm{C}}, \mathbf{e}_{\mathrm{C}}} \frac{1}{2} \|\boldsymbol{y}_{\mathrm{C}} - \boldsymbol{\Phi}_{\mathcal{M}} \boldsymbol{x}_{\mathrm{C}} - \mathbf{e}_{\mathrm{C}}\|_2^2 + a_{\mathrm{C}} \|\boldsymbol{x}_{\mathrm{C}}\|_1 + b_{\mathrm{C}} \sum_{k=1}^{K_{\mathrm{C}}} \|\mathbf{e}_{\mathrm{C,k}}\|_2 \quad (7)$$

where $\|\boldsymbol{x}\|_1 = \sum_n |x_n|$ is the ℓ_1 norm of $\boldsymbol{x}$, $\mathbf{e}_{\mathrm{C},k}$ corresponds to the kth time-series of $\mathbf{e}_{\mathrm{C}}$ associated with the kth continuous parameter with $\mathrm{k} = 1, ..., K_{\mathrm{C}}$, a_{C} and b_{C} are regularization parameters that control the level of sparsity of the coefficient vector $\boldsymbol{x}_{\mathrm{C}}$ and the anomaly signal $\mathbf{e}_{\mathrm{C}}$, respectively. The dictionary $\Phi_{\mathcal{M}}$ is composed of the continuous atoms $\boldsymbol{\phi}_{\mathrm{C},l}$ for $l \in \mathcal{M}$. The subset $\mathcal{M}$ allows the selection of representative continuous atoms and the preservation of relationships between discrete and continuous parameters and the detection of multivariate anomalies. For that reason the sparse subproblem in recovering discrete values is first dealt with, rather than the other way around. Note that (7) considers two distinct sparsity constraints for the coefficient vector $\boldsymbol{x}_{\mathrm{C}}$ and the anomaly signal $\mathbf{e}_{\mathrm{C}}$. This formulation reflects the fact that a nominal continuous signal can be well approximated by a linear combination of *few* atoms of the dictionary (sparsity of $\boldsymbol{x}_{\mathrm{C}}$) and that anomalies are *rare* and affect few parameters at the same time (sparsity of $\mathbf{e}_{\mathrm{C}}$).

Problem (7) can be solved with the alternating direction method of multipliers (ADMM) [17] by adding an auxiliary variable $\mathbf{z}$

$$\min_{\boldsymbol{x}_{\mathrm{C}}, \mathbf{e}_{\mathrm{C}}, \mathbf{z}} \frac{1}{2} \|\boldsymbol{y}_{\mathrm{C}} - \boldsymbol{\Phi}_{\mathcal{M}} \boldsymbol{x}_{\mathrm{C}} - \mathbf{e}_{\mathrm{C}}\|_2^2 + a_{\mathrm{C}} \|\mathbf{z}\|_1 + b_{\mathrm{C}} \sum_{k=1}^{K_{\mathrm{C}}} \|\mathbf{e}_{\mathrm{C,k}}\|_2 \quad (8)$$

and the constraint $\mathbf{z} = \boldsymbol{x}_{\mathrm{C}}$. Note that, contrary to Problem (7), the first and second terms of (8) are decoupled, which allows an easier estimation of the vector $\boldsymbol{x}_{\mathrm{C}}$. The ADMM algorithm associated with (8) minimizes the following augmented Lagrangian

$$\mathcal{L}_A(\boldsymbol{x}_{\mathrm{C}}, \mathbf{z}, \mathbf{e}_{\mathrm{C}}, \mathbf{m}, \mu) = \frac{1}{2} \|\boldsymbol{y}_{\mathrm{C}} - \boldsymbol{\Phi}_{\mathcal{M}} \boldsymbol{x}_{\mathrm{C}} - \mathbf{e}_{\mathrm{C}}\|_2^2 + a_{\mathrm{C}} \|\mathbf{z}\|_1$$
$$+ b_{\mathrm{C}} \sum_{k=1}^{K_{\mathrm{C}}} \|\mathbf{e}_{\mathrm{C,k}}\|_2 + \mathbf{m}_{\mathrm{C}}^T (\mathbf{z} - \boldsymbol{x}_{\mathrm{C}}) + \frac{\mu_{\mathrm{C}}}{2} \|\mathbf{z} - \boldsymbol{x}_{\mathrm{C}}\|_2^2 \quad (9)$$

where $\mathbf{m}_{\mathrm{C}}$ is a Lagrange multiplier and μ_{C} is a regularization parameter controlling the level of deviation between $\boldsymbol{z}$ and $\boldsymbol{x}_{\mathrm{C}}$. The ADMM algorithm is iterative and alternatively estimates $\boldsymbol{x}_{\mathrm{C}}$, $\mathbf{z}$, $\mathbf{e}_{\mathrm{C}}$ and $\mathbf{m}_{\mathrm{C}}$. More details about the update equations of the different variables at the kth iteration are provided [11].

C. Anomaly score

The proposed method defines an anomaly score assigned to each test signal $\boldsymbol{y}$ such as

$$a(\boldsymbol{y}) = \begin{cases} -1 & \text{if } \mathcal{M} = \emptyset \\ \| \ \widehat{\boldsymbol{e}}_{\mathrm{C}} \ \|_2 & \text{otherwise.} \end{cases} \quad (10)$$

The anomaly score is used to detect possible anomaly in $\boldsymbol{y}$ using the following rule

$$\text{anomaly detected if } \begin{cases} a(\boldsymbol{y}) = -1 & \text{(Discrete AD)} \\ a(\boldsymbol{y}) > S_{\mathrm{PFA}} & \text{(Continuous AD)} \end{cases} \quad (11)$$

where $S_{\mathrm{PFA}} > 0$ is a threshold depending on the probability of false alarm of the anomaly detector (which has to be adjusted by the user). This threshold can be determined using receiver operating characteristic (ROC) curves if a ground-truth is available.

D. Shift Invariant Option

The proposed method has a shift-invariant (SI) option allowing new discrete atoms to be created by shifting of τ lags each atoms of $\boldsymbol{\Phi}_{\mathrm{D}}$, with $\tau \in \{-\tau_{\max}, -(\tau_{\max} - 1), ..., -1, 0, 1, ..., \tau_{\max} - 1, \tau_{\max}\}$. This shift invariant option was proposed in some recent works including [18]. The maximum shift $\tau_{\max}$ has to be fixed by the user, in order to have an acceptable computation cost and a good detection performance. Experimental results showed the usefulness of the SI option notably for discrete time-series, which will be illustrated in our experiments.

III. Experimental Results

The proposed multivariate Anomaly Detection method based on a sparse decomposition and DICTionary learning [19] was evaluated on a representative anomaly dataset composed of actual anomalies that occurred on several operated satellites and compared to three state-of-the-art methods. In the first experiment, we consider a simple dataset composed of $K_{\mathrm{D}} = 3$ discrete and $K_{\mathrm{C}} = 7$ continuous parameters with available ground-truth. The dictionary was learnt using two months of telemetry describing nominal behaviours of parameters, which represents 30000 training signals obtained after applying the preprocessing with the following parameters $\delta = 5$ and $w = 50$. Unfortunately, there is no dictionary learning method appropriate to mixed discrete and continuous data. Our dictionary was built as follows: 1) set a number of desired atoms ($L = 2000$ in our simulations), 2) initialize a continuous dictionary with L continuous training signals randomly selected in the training set, 3) each continuous training signal of the learning dataset is decomposed into the initial dictionary by solving (1) with $\boldsymbol{e} = \mathbf{0}$, which is the well-known Lasso problem [20], 4) select the L training

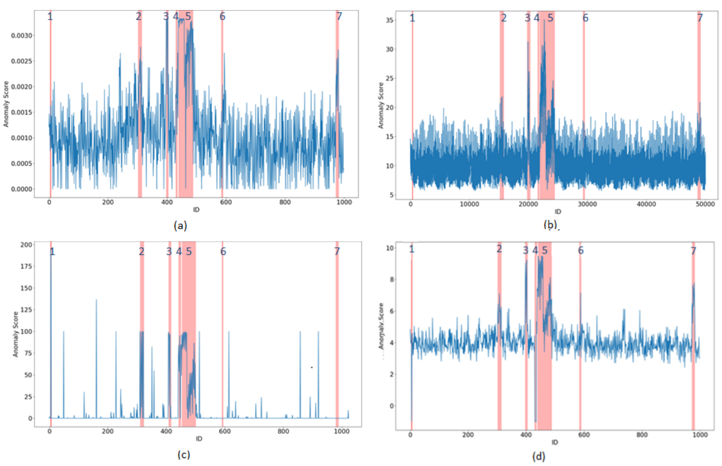

Fig. 3. Anomaly scores returned by OCSVM (a), MPPCAD (b), NOSTRADAMUS (c) and ADDICT (d) for each test signals of the anomaly dataset with ground-truth marked by blue background.

signals with the biggest residuals $\| \boldsymbol{y}_C - \boldsymbol{\Phi}\boldsymbol{x}_C \|$. This process was repeated 100 times and the L training signals most often selected during the different iterations were selected as atoms of the dictionary. The ADDICT algorithm was evaluated on a representative anomaly dataset composed of 1000 test signals including 90 known anomalies. Note that the anomalies were located in 7 time periods displayed in Fig. 1 (in the red boxes). Three state-of-the-art methods were evaluated on this dataset

- the NOSTRADAMUS algorithm [3]. It is a univariate algorithm based on the one-class support vector machine (OC-SVM) algorithm with an appropriate preprocessing
- the multivariate OC-SVM algorithm [21] with the ADDICT preprocessing
- the mixture of probabilistic principal component analysers and categorical distribution (MPPCAD) algorithm [9], which is a multivariate anomaly detection method based on probabilistic clustering and dimensionality reduction. The idea behind MPCCAD is to approximate the joint distribution of the continuous variables by a mixture of Gaussian distributions and the joint distribution of discrete variables by a mixture of categorical distributions.

Fig. 3 shows the anomaly scores (in blue) returned by OC-SVM (a), MPPCAD (b), NOSTRADAMUS (c) and ADDICT (d) for each signal of the anomaly dataset with ground-truth marked by red backgrounds and hyperparameters of baseline approaches are tuned by cross validation. Anomaly periods (red backgrounds) are numbered and displayed in Fig. 1. Some anomalies are well detected by all the methods. This includes anomalies $\#2$, $\#4$ and $\#5$ for which high scores are returned for each method. On the other hand, some anomalies such as the ones affecting discrete parameters (see anomalies $\#1$ and $\#3$) are more difficult to detect. These anomalies are not detected by OC-SVM and MPPCAD. They are detected by ADDICT when the shift-invariance option is activated, which shows the importance of this option. Finally, note that the multivariate anomaly $\#7$ is well detected by all the multivariate methods, namely OC-SVM, MPPCAD and ADDICT. This anomaly is not detected by NOSTRADAMUS which is a univariate AD method handling the different times series separately.

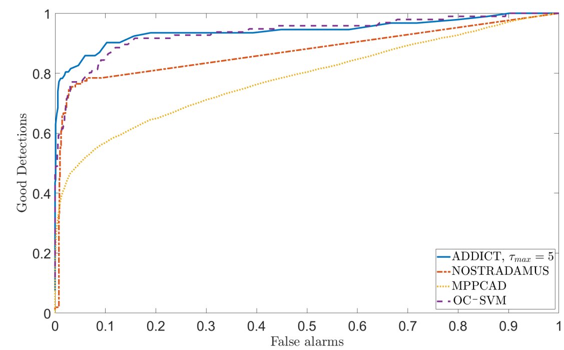

Fig. 4. ROC curves of OC-SVM, NOSTRADAMUS, MPPCAD and ADDICT for the anomaly dataset.

TABLE I. VALUES OF P_{D} AND P_{FA} FOR OCSVM, MPPCAD, NOSTRADAMUS AND ADDICT.

Method	Threshold	P_{D}	P_{FA}
MPPCAD	13	55%	8%
OC-SVM	0.0016	80%	7%
NOSTRADAMUS	29	77.26%	6%
ADDICT ($\tau_{\max} = 0$)	5.2	80%	6.16%
ADDICT ($\tau_{\max} = 5$)	5.4	85.56%	3.3%

In order to quantify the detection performance of each method, Fig. 4 displays the receiver operational characteristics (ROCs) expressing the probability of detection P_{D} as a function of the probability of false alarm P_{FA}. Quantitative results are are also summarized in Table I, which reports the values of P_{D} and P_{FA} satisfying the best compromise for spacecraft health monitoring ($P_{\text{FA}} < 10\%$), for which the control of false alarms is a major issue (since all false anomalies are checked by a human operator). In view of these results, NOSTRADAMUS, OC-SVM and ADDICT are more competitive with almost 80% of detections for less than 10% of false alarms. MPPCAD returns few false alarms but an important proportion of anomalies from the anomaly dataset are not detected. These non-detections can be partly explained by the MPPCAD preprocessing which segments telemetry time-series into windows of size $w = 1$. Indeed, the detection of anomalies $\#1$, $\#4$ or $\#6$ for example, clearly requires longer time windows to be considered.

IV. CONCLUSION

This paper presented a new anomaly detection method for mixed (discrete and continuous) telemetry time-series based on a sparse decomposition on a dictionary of nominal patterns. The first results showed that the proposed method can detect all kinds of anomalies (i.e., univariate and multivariate anomalies) and is competitive with respect to the state-of-the-art. In addition, our experiments showed the importance of the shift-invariant option that allowed a significant reduction of the false alarm rate. For future works, it is important to evaluate the method in an operating context including hundreds to thousands telemetry parameters. Furthermore, we would like to investigate an online extension of the ADDICT algorithm, allowing user feedback to be integrated, in order to update the model and improve detection performance.

ACKNOWLEDGMENT

The authors would like to thank Pierre-Baptiste Lambert from CNES and Clémentine Barreyre from Airbus Defence and Space for fruitful discussions about anomaly detection in spacecraft telemetry. This work was supported by CNES and Airbus Defence and Space.

References

[1] V. Chandola, A. Banerjee, and V. Kumar, "Anomaly detection: A survey," *ACM Computing Survey*, vol. 43, no. 3, Jul. 2009.

[2] ——, "Anomaly detection for discrete sequences: A survey," *ACM Computing Survey*, vol. 24, no. 5, May 2012.

[3] S. Fuertes, G. Picard, J.-Y. Tourneret, L. Lofti, A. Ferrari, and C. Richard, "Improving spacecraft health monitoring with automatic anomaly detection techniques," in *Proc. Int. Conf. Space Operations (SpaceOps'2016)*, Daejeon, South Korea, May 2016.

[4] I. Verzola, A. Donati, J.-A. M. Heras, M. Schubert, and L. Somodi, "Project sybil: A novelty detection system for human spaceflight operations," in *Proc. Int. Conf. Space Operations (SpaceOps'2016)*, Daejeon, South Korea, May 2016.

[5] J.-A. Martínez-Heras, A. Donati, M. Kirksch, and F. Schmidt, "New telemetry monitoring paradigm with novelty detection," in *Proc. Int. Conf. Space Operations (SpaceOps'2012)*, Stockholm, Sweden, Jun. 2012.

[6] C. O'Meara, L. Schlag, L. Faltenbacher, and M. Wickler, "ATHMoS: Automated telemetry health monitoring system at GSOC using outlier detection and supervised machine learning," in *Proc. Int. Conf. Space Operations (SpaceOps'2016)*, Daejeon, South Korea, May 2016.

[7] C. O'Meara, L. Schlag, and M. Wickler, "Applications of deep learning neural networks to satellite telemetry monitoring," in *Proc. Int. Conf. Space Operations (SpaceOps'2018)*, Marseille, France, Jun. 2018.

[8] N. Takeishi and T. Yairi, "Anomaly detection from multivariate times-series with sparse representation," in *Proc. IEEE Int. Conf. Syst. Man and Cybernetics*, San Siego, CA, USA, Oct. 2014.

[9] T. Yairi, N. Takeishi, T. Oda, Y. Nakajima, N. Nishimura, and N. Takata, "A data-driven health monitoring method for satellite housekeeping data based on probabilistic clustering and dimensionality reduction," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 53, no. 3, pp. 1384–1401, Jun. 2017.

[10] C. Barreyre, "Statistiques en grande dimension pour la détection d'anomalies dans les données fonctionnelles issues des satellites," Ph.D. dissertation, Université de Toulouse, Toulouse, France, 05 2018.

[11] A. Adler, M. Elad, Y. Hel-Or, and E. Rivlin, "Sparse coding with anomaly detection," *J. Signal Process. Syst.*, vol. 79, no. 2, pp. 179–188, 2015.

[12] M. Elad and A. Aharon, "Image denoising via sparse and redundant representation over learned dictionaries," *IEEE Trans. Image Process.*, vol. 14, no. 12, pp. 3736–3745, Dec. 2006.

[13] Y. Chen, N. M. Nasrabadi, and T. D. Tran, "Hyperspectral image classification using dictionary-based sparse representation," *IEEE Trans. Geosci. Remote Sens.*, vol. 49, no. 10, pp. 3973–3985, May. 2011.

[14] J. Wright, A. Y. Yang, A. Ganesh, S. S. Sastry, and Y. Ma, "Robust face recognition via sparse representation," *IEEE Trans. Patt. Anal. Mach. Intell.*, vol. 31, no. 3, pp. 1–18, Feb. 2009.

[15] Y. Xu, Z. Wu, J. Li, A. Plaza, and Z. Wei, "Anomaly detection in hyperspectral images based on low-rank and sparse representation," *IEEE Trans.Geosci. Remote Sens.*, vol. 54, no. 4, pp. 1990–2000, Apr. 2016.

[16] S. Biswas and R. Venkatesh, "Sparse representation-based anomaly detection with enhanced local dictionaries," in *Proc. IEEE Int. Conf. Image Process.*, Paris, France, Oct. 2014.

[17] S. Boyd, N. Parikh, E. Chu, B. Peleato, and J. Eckstein, "Distributed optimization and statistical learning via alternating direction method of multipliers," *Foundations and Trends in Machine Learning*, vol. 3, no. 1, pp. 1–122, 2010.

[18] Q. Barthélemy, C. Gouy-Pailler, Y. Isaac, A. Souloumiac, A. Larue, and J. Mars, "Multivariate temporal dictionary learning for EEG," *J. Neurosci. Methods*, vol. 215, no. 1, pp. 19–28, 2013.

[19] B. Pilastre, S. d'Escrivan, L. Boussouf, and J.-Y. Tourneret, "Anomaly detection in mixed telemetry data using a sparse representation and dictionary learning," *Signal Process.*, submitted, May 2019.

[20] R. Tibshirani, "Regression shrinkage and selection via the LASSO," *J. Roy. Statist. Soc. Series B (Methodological)*, vol. 58, no. 1, pp. 267–288, 1996.

[21] B. Schölkopf, J. C. Platt, J. Shawe-Taylor, A. J. Smola, and R. C. Williamson, "Estimating the support of a high-dimensional distribution," *Neural Computation*, vol. 3, no. 7, pp. 1443–1471, 2001.